Adaptive Coded Aperture Design by Motion Estimation using Convolutional Sparse Coding in Compressive Spectral Video Sensing

N. Diaz[1], C. Noriega-Wandurraga[2], A. Basarab[3], J.-Y. Tourneret[4] and H. Arguello[5]
[1]Department of Electrical Engineering, Universidad Industrial de Santander (UIS), 680002 Bucaramanga, Colombia
[2]Department of Mathematics, UIS, 680002 Bucaramanga, Colombia
[3]University of Toulouse, IRIT, CNRS UMR 5505, F-31062 Toulouse, France, Email: adrian.basarab@irit.fr
[4]University of Toulouse, IRIT/INP-ENSEEIHT/TéSA, 31071 Toulouse, France, Email: jean-yves.tourneret@enseeiht.fr
[5]Department of Computer Science, UIS, 680002 Bucaramanga, Colombia, Email: henarfu@uis.edu.co

***Abstract*—This paper proposes a new motion estimation method based on convolutional sparse coding to adaptively design the colored-coded apertures in static and dynamic spectral videos. The motion in a spectral video is estimated from a low-resolution reconstruction of the datacube by training a convolutional dictionary per spectral band and solving a minimization problem. Simulations show improvements in terms of peak signal-to-noise ratio (of up to 2 dB) of the reconstructed videos by using the proposed approach, compared with state-of-art non-adaptive coded apertures.**

***Index Terms*—Compressive spectral video, motion estimation, adaptive imaging, convolutional sparse coding.**

I. Introduction

Compressive spectral imaging (CSI) systems drastically reduce the amount of acquired spectral data by capturing datacube projections in order to reconstruct the underlying image. One outstanding CSI architecture is the colored-coded aperture snapshot spectral imager (C-CASSI) [1]. C-CASSI is a snapshot system that captures compressive projections of the datacube along time. C-CASSI uses three optical components, the colored-coded aperture (CCA), the dispersive element and the focal plane array (FPA). Specifically, the incoming light of the scene is spatially and spectrally modulated by the colored-coded aperture and spectrally smeared by the dispersive element before it impinges on the FPA [1]. The underlying spectral scene is reconstructed by solving a convex optimization problem [2], [3]. Unlike traditional CASSI that uses block-unblock coded apertures (which block or transmit the full spectral signature in a given pixel [4]), the colored-coded apertures filter the incoming light per pixel. C-CASSI has been used successfully for capturing static scenes. Extending C-CASSI to dynamic scenes is clearly interesting.

Recently, a variation of C-CASSI based on compressive spectral video sensing (CSVS), called video-C-CASSI, has been proposed [5]. Most CSVS systems rely on coding and dispersion of the incoming light towards the camera sensor [6]–[10]. In addition, video-CASSI improves flexibility in the encoding of the dynamic scenes by using an array of optical filters. Moreover, video-C-CASSI employs uniform sampling along the frames, estimates the motion between successive frames using optical flow and uses a regularization term to reduce the errors introduced by the motion of the scene [5]. While optical flow is a well-established algorithm for motion estimation, recent works showed that motion regularization based on patch-based or convolutional learned dictionaries generally improves motion estimation (see, e.g., [11]–[13]). However, the interest of these approaches for the design of the colored-coded aperture of dynamic scenes has never been explored, which is the main objective of this work.

A traditional coded aperture design of C-CASSI is performed in a non-adaptive manner, i.e., the coded aperture is designed independently from the scene [1], [5]. Recently, a few works proposed adaptive designs in C-CASSI [14], and CASSI [15] using static scenes. The main objective of this paper is to use convolutional sparse coding (CSC), a translation-invariant image representation [16] to compute the motion between successive frames of a video, and to consequently design coded apertures adaptively by separating the sampling of static and dynamic scenes. The contribution of this work is to design coded apertures of C-CASSI adaptively to capture compressive measurements for spectral video. The proposed method estimates the motion between successive frames by minimizing a cost function formed by a data attachment term penalized by a spatial regularization promoting smoothness and a sparse regularization using a convolutional dictionary for motion fields. Simulations results show that the proposed adaptive video-C-CASSI outperforms approaches of the state-of-art such as random C-CASSI and blue-noise C-CASSI.

II. Video C-CASSI

Video C-CASSI acquires dynamic scenes at a particular frame rate. It is composed of an objective lens, a temporal colored-coded aperture (T-CCA), a relay lens, a dispersive element, and a focal plane array (FPA) or detector. The discrete measurements for the dth frame on the detector are

$$Y^d_{i,j} = \sum_{\ell=0}^{L-1} \mathbf{F}^d_{i,(j+\ell),\ell} T^d_{i,(j+\ell),\ell} + \omega_{i,j} \tag{1}$$

where $Y^d_{i,j}$ are the elements of a matrix $\mathbf{Y}^d \in \mathbb{R}^{N\times(M+L-1)}$ containing the measurements of the dth frame, $M \times (N + L - 1)$ is the dimension of the detector (note that the width of the compressive measurements is higher than the height due to the dispersion of the prism), and $\omega_{i,j}$ is the Gaussian noise of the sensing system at position (i, j). The acquisition process can be compactly written in matrix form as $\mathbf{y}^d = \mathbf{H}^d\mathbf{f}^d + \boldsymbol{\omega}^d$ where $\mathbf{y}^d \in \mathbb{R}^{N(M+L-1)}$ is a vector containing the compressive measurements of $\mathbf{Y}^d$ and $\mathbf{f}^d \in \mathbb{R}^{NML}$ is the vectorized datacube $\mathbf{F}^d \in \mathbb{R}^{N\times M\times L}$ for frame $\#d$, $\boldsymbol{\omega}^d$ is the corresponding vectorized Gaussian noise and $\mathbf{H}^d$ is the dth C-CASSI sensing matrix whose structure was defined in [1]. Note that $\mathbf{H}^d$ models the physical phenomenon of the coded aperture and the shifting produced by the dispersive element. More details about the structure of the matrix $\mathbf{H}^d$ can be found in [1].

III. Coded aperture design in spectral videos using motion estimation.

A. Motion Estimation

Pairwise video motion estimation (VME) uses two consecutive frames $\mathbf{F}^{d-1}_H$ and $\mathbf{F}^d_H$ (of $\mathbb{R}^{M\times N\times L}$) from a spectral video acquired at time instants $d-1$ and d. Denote as $\mathbf{S}^d_{(\ell,x)}$ and $\mathbf{S}^d_{(\ell,y)} \in \mathbb{R}^{M\times N\times L}$ the video motions for the frame d along the x and y axes. The proposed VME method is inspired by the works conducted in [11]. It minimizes a function composed of a data fidelity term E_{data}, penalized by two regularization terms E_{spatial} and E_{sparse}

$$\underset{\mathbf{X},\mathbf{S}^d}{\operatorname{argmin}}\left\{E_{\text{data}}(\mathbf{S}^d,\mathbf{F}^d_H,\mathbf{F}^{d-1}_H)+\lambda_s E_{\text{spatial}}(\mathbf{S}^d)+\lambda_p E_{\text{sparse}}(\mathbf{S}^d,\mathbf{X})\right\} \tag{2}$$

where $(\lambda_p, \lambda_s) \in \mathbb{R}^2$ are two regularization parameters (balancing the influence of the data fidelity term and the regularizations) and $\mathbf{S}^d = \mathbf{S}^d_{(\ell,x)}$ or $\mathbf{S}^d = \mathbf{S}^d_{(\ell,y)}$. Note that the displacement vectors components along x and y are estimated independently for simplicity. The first regularization term promotes smooth variations in the video motion field by using a standard total variation function, i.e., $E_{\text{spatial}}(\mathbf{S}^d) = \|\nabla\mathbf{S}^d\|^2_2$, where ∇ is the gradient operator and $\|.\|^2_2$ is the squared ℓ_2 norm. The other terms in (2) are defined below.

B. Data fidelity term

Optical flow assumes brightness constancy and temporal consistency leading to the following optical flow equation

$$\partial_t\mathbf{f}^d_H + \nabla\mathbf{f}^T_H\mathbf{s}^d = 0 \tag{3}$$

where $\mathbf{s}^d \in \mathbb{R}^{NM}$ represents the flow field such that $\mathbf{s}^d_\ell$ is the vectorized video motion $\mathbf{S}_\ell$, $\partial_t\mathbf{f}^d_H$ denotes the temporal derivative and $\nabla\mathbf{f}^T_H$ is the spatial gradient of the brightness. The data fidelity term resulting from optical flow is

$$E_{\text{data}}(\mathbf{s}^d,\mathbf{f}^d_H,\mathbf{f}^{d-1}_H) = \left\|\partial_t\mathbf{f}^d_H + \nabla\mathbf{f}^T_H\mathbf{s}^d\right\|^2_2 \tag{4}$$

where $\|.\|^2_2$ is the squared ℓ_2 norm.

C. Sparse regularization

The second regularization term promotes sparsity of the motion vectors in a dictionary of representative motions. It decomposes the video motion $\mathbf{S}^d$ as a convolution between V sparse coefficient maps $\mathbf{X}_v$ and a set of V filters $\mathbf{G}_v$, i.e.,

$$E_{\text{sparse}}(\mathbf{S}^d,\mathbf{X}) = \left\|\mathbf{S}^d - \sum_{v=1}^{V}\mathbf{G}_v * \mathbf{X}_v\right\|^2_2 \tag{5}$$

where $*$ denotes convolution. This regularization was used successfully for ultrasound images [13]. One of the objectives of this paper is to analyze its interest for multi-temporal hyperspectral images.

D. Adaptive coded aperture scheme

Fig. 1 summarizes the four steps proposed to design the video adaptive colored-coded apertures (VA-CCA). Specifically, the following iterative process is repeated for all the frames: (1) sample the datacube to capture the compressive measurements of a pair of frames, (2) reconstruct the datacube by solving an inverse problem to define some prior information [14], (3) estimate the video motions using optical flow, (4) design the colored-coded apertures by thresholding the motion estimates resulting from (3). Note that step (3) requires the computation of a convolutional dictionary (step (3a)) (which uses the training video motion $\widetilde{\mathbf{S}}$) and of coefficient maps ((step (3b)) using test video motions $\mathbf{S}_t$.

E. Adaptive coded aperture design algorithm

Algorithm 1 designs the coded apertures adaptively for a compressive spectral video sequence. The algorithm uses two sets of compressive measurements and a sequence of motion fields estimated from a spectral video. It also requires to choose the hyperparameters λ_s, λ_p that balance the spatial and sparse regularization terms, respectively. Moreover, the alternating direction method of multipliers (ADMM) [17] used in this work requires to set the parameter ρ for the Lagrangian, and λ to balance the sparsity of the coefficients maps. Note that a low-resolution reconstruction of the datacube is necessary to obtain the video sequence appearing in lines 6 and 12 of the algorithm. This reconstruction is obtained by solving the following optimization problem $\hat{\mathbf{f}}^d_L = \boldsymbol{\Psi}^{-1}_L(\operatorname{argmin}_{\boldsymbol{\theta}_L}\|\mathbf{y}^d - \mathbf{H}^d_L\boldsymbol{\Psi}^d_L\boldsymbol{\theta}^d_L\|^2_2 + \tau\|\boldsymbol{\theta}^d_L\|_1)$, where $\mathbf{y}$ contains the compressive measurements, $\boldsymbol{\Psi}^{-1}_L$ denotes the low-resolution representation basis, $\boldsymbol{\theta}^d_L$ represents the sparse signal, $\mathbf{H}^d_L$ corresponds to the measurement matrix, $\|.\|_1$ is the ℓ_1-norm, and τ is a regularization parameter. Line 2 computes the convolutional dictionary by assuming that each band in a specific spectral frame of the video can be represented as a set of V filters $\mathbf{G}_v$ convolved with a set of sparse coefficient maps $\mathbf{X}_v$, i.e., $\mathbf{S}_d \approx \sum_{v=1}^{V}\mathbf{G}_v * \mathbf{X}_v$.

F. Dictionary learning

The determination of a convolutional dictionary satisfying $\mathbf{S}_d \approx \sum_{v=1}^{V}\mathbf{G}_v * \mathbf{X}_v$ can be achieved using several off-the-shelf algorithms [18]. In this paper, the dictionary learning

step is performed by solving the following problem (where $\widetilde{S}_d$ denotes the training video sequence which was obtained using Horn-Schunck optical flow estimation)

$$\underset{\mathbf{G}_v,\mathbf{X}_{d,v}}{\operatorname{argmin}} \frac{1}{2}\sum_d \left\| \sum_v \mathbf{X}_{d,v} * \mathbf{G}_v - \widetilde{\mathbf{S}}_d \right\|_2^2 + \lambda \sum_{v=1}^{V}\sum_d \|\mathbf{X}_{d,v}\|_1 \quad \text{s.t.} \quad \|\mathbf{G}_v\| = 1 \ \forall v = 1, ..., V. \tag{6}$$

The minimization of (6) can be handled efficiently using the ADMM.

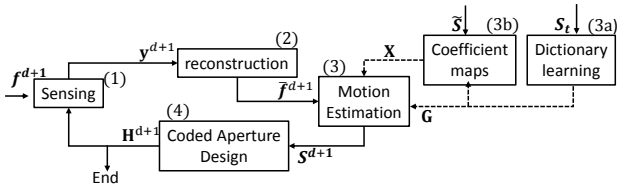


Fig. 1. Flowchart of the proposed adaptive coded aperture design for VME.

Algorithm 1 Adaptive coded aperture design for compressive spectral video using motion estimation.

Input: $\lambda_s, \lambda_p, K, D, \lambda, \rho,\ \widetilde{\mathbf{S}}, \mathbf{S}_t$: Training/test video motions
Output: $\mathbf{S}_\ell^d$

1: **function** CODED APERTURE DESIGN USING VIDEO MOTION ESTIMATION $(\mathbf{y}^0, \mathbf{y}^1, \lambda_s, \lambda_p, K, J, \lambda, \rho, \widetilde{\mathbf{S}}, \mathbf{S}_t)$
2: $\mathbf{G}_v \leftarrow$ Computes the dictionary by solving (6)
3: $\mathbf{X}_v \leftarrow$ Computes the coefficient maps by solving (7)
4: $\mathbf{y}^0 \leftarrow \mathbf{H}^0\mathbf{f}$ ▷ First snapshot
5: $\hat{\mathbf{f}}_L^0 \leftarrow \mathbf{\Psi}_L^{-1}(\operatorname{argmin}_{\boldsymbol{\theta}_L} \|\mathbf{y}^0 - \mathbf{H}_L^0\mathbf{\Psi}_L^d\boldsymbol{\theta}_L^d\|_2^2 + \tau\|\boldsymbol{\theta}_L^d\|_1)$
6: ▷ Low-resolution
7: $\hat{\mathbf{f}}_H^0 \leftarrow \mathbf{P}(\hat{\mathbf{f}}_L^0)$ ▷ Interpolation
8: $\hat{\mathbf{F}}_H^0 \leftarrow rearrange(\hat{\mathbf{f}}_H^0)$ ▷ Rearrange
9: **for** $k \leftarrow 1, K$ **do**
10: **for** $d \leftarrow 1, D$ **do**
11: $\hat{\mathbf{f}}_L^d \leftarrow \mathbf{\Psi}_L^{-1}(\operatorname{argmin}_{\boldsymbol{\theta}_L} \|\mathbf{y}^d - \mathbf{H}_L^d\mathbf{\Psi}_L^d\boldsymbol{\theta}_L^d\|_2^2 + \tau\|\boldsymbol{\theta}_L^d\|_1)$
12: ▷ Low-resolution
13: $\hat{\mathbf{f}}_H^d \leftarrow \mathbf{P}(\hat{\mathbf{f}}_L^d)$ ▷ Interpolation
14: $\hat{\mathbf{F}}_H^d \leftarrow rearrange(\hat{\mathbf{f}}_H^d)$ ▷ Rearrange
15: **for** $\ell \leftarrow 1, L$ **do**
16: $\operatorname{argmin}_{\mathbf{S}_\ell^d}\{E_{\text{data}}(\hat{\mathbf{F}}_H^{d-1}, \hat{\mathbf{F}}_H^d, \mathbf{S}_\ell^{d-1}) + \lambda_s\|\nabla\mathbf{S}_\ell^{d-1}\|_2^2 + \lambda_p(k)\|\mathbf{S}_\ell^{d-1} - \sum_v \mathbf{G}_v * \mathbf{X}_v\|_2^2\}$
s.t. $\|\mathbf{G}_v\| = 1\ \forall v$ ▷ Video motion estimation
17: $\mathbf{Q}_\ell^d \leftarrow (\mathbf{S}_\ell^{d-1}, \mathbf{S}_\ell^d)$ ▷ Thresholding motion
18: $\mathbf{q}_\ell^d \leftarrow vec(\mathbf{Q}_\ell^d)$ ▷ Vectorized motion areas
19: $\mathbf{r}_\ell^d \leftarrow \mathbf{q}_\ell^d \odot \mathbf{b}_\ell^d + (\mathbf{1} - \mathbf{q}_\ell^d) \odot \hat{\mathbf{b}}_\ell^d$ ▷ Next code
20: $\mathbf{H}_\ell^d \leftarrow rearrange(\mathbf{r}_\ell^d)$ ▷ Rearrange
21: $\mathbf{y}^d \leftarrow \mathbf{H}^d\mathbf{f}$ ▷ Next snapshot
22: $\hat{\mathbf{f}} \leftarrow \mathbf{\Psi}^{-1}(\operatorname{argmin}_{\boldsymbol{\theta}} \|\mathbf{y} - \mathbf{H}\mathbf{\Psi}\boldsymbol{\theta}\|_2^2 + \tau\|\boldsymbol{\theta}\|_1)$
23: **return** $\mathbf{S}_\ell^d$ ▷ (Estimated motion field)

G. Sparse coding

Once the dictionary $\mathbf{G}_v$ has been determined, the coefficient maps of a sequence of test images denoted as $\mathbf{S}_t^d$ are obtained by solving the following optimization problem (see line 3 of Algorithm 1)

$$\underset{X_v}{\operatorname{argmin}} \frac{1}{2}\left\| \sum_{v=1}^{V} \mathbf{X}_v * \mathbf{G}_v - \mathbf{S}_t^d \right\|_2^2 + \lambda \sum_{v=1}^{V} \|\mathbf{X}_v\|_1 \tag{7}$$

which can again be done using the ADMM algorithm.

H. Video reconstruction

In order to compute the adaptive coded apertures, some prior information is required. We propose to use a low resolution reconstruction of the image $\hat{\mathbf{f}}_L^d$ (see line 11 of Algorithm 1), where $\mathbf{H}_L^d$ is a decimated sensing matrix $\mathbf{H}_L^d = \mathbf{H}^d\mathbf{D}$, $\mathbf{D}$ is another decimation matrix such that $\hat{\mathbf{f}}_L^d = \mathbf{D}\hat{\mathbf{f}}^d$, the representation basis is denoted by $\mathbf{\Psi}_L^{-1}$ and the vectorized sparse signal is $\boldsymbol{\theta}_L^d$. An interpolation is used in Line 13 of Algorithm 1 to create a high resolution image $\hat{\mathbf{f}}_H^d$ from a low resolution reconstruction $\hat{\mathbf{f}}_L^d$, where $\mathbf{P}$ is a bilinear interpolator. The interpolated datacube $\hat{\mathbf{f}}_H^d$ is rearranged in Line 14 to obtain the datacube of the dth frame $\hat{\mathbf{F}}_H^d$.

The VME is reported in Line 16. In line 17 of the algorithm, a thresholding using a binary mask $\mathbf{Q}_\ell^d$ divides each pixel in background and target areas using a pair of motions $\mathbf{S}_\ell^{d-1}$, and $\mathbf{S}_\ell^d$. The matrix $\mathbf{Q}_\ell^d$ is vectorized as $\mathbf{q}_\ell^d$ in line 18 and the resulting coded aperture $\mathbf{r}_\ell^d$ is computed in line 19. Note that the coded aperture depends on two blue noise code apertures $\hat{\mathbf{b}}_\ell^d$ and $\mathbf{b}_\ell^d$ [19]. One blue noise coded aperture corresponds to the background area (complement of matrix $\mathbf{q}_\ell^d$, $\mathbf{1} - \mathbf{q}_\ell^d$) and a moving blue noise coded aperture $\mathbf{b}_\ell^d$ is attributed to the mobile target (the subset of the scene $\mathbf{q}_\ell^d$). For an example of designed codes, the reader is invited to look at Fig. 2 in Section IV. In line 20 the coded aperture is reorganized leading to the matrix $\mathbf{H}_\ell^d$. In the following step (line 21) the compressive measurements $\mathbf{y}_d$ are obtained by sampling with the adaptive coded aperture $\mathbf{H}_\ell^d$. The last step in line 22 reconstructs the spectral video with improved quality due to the designed patterns.

IV. Simulation results

In order to validate the performance of the proposed coded aperture design, a set of C-CASSI video measurements was simulated using the model (1). These measurements were constructed using a real test spectral source acquired in the Optics Lab of the High Dimensional Signal Processing (HDSP) research group at Universidad Industrial de Santander with a CCD camera using wavelength steps of 10 nm. The resulting discrete source $\mathbf{F}$ used in simulations has 12 frames of 128×128 pixels and $L = 10$ spectral bands ranging from 400 nm to 500 nm. Given the compressive projections, the compressive sensing algorithm GPSR (Gradient Projection for Sparse Reconstruction) was used to recover the data [3]. The 4D sparse representation basis used in this experiment was the Kronecker product between a 2D-Wavelet Symmlet 8 basis for the spatial dimensions denoted as $\mathbf{\Psi^{2D}}$, a 1D-Discrete Cosine basis (DCT) for the spectral dimension denoted as $\mathbf{W}$ and a 1D-DCT basis for the temporal dimension denoted as $\mathbf{U}$ [20].

Algorithm 1 was used to design the video adaptive colored-coded apertures (VA-CCA) in an adaptive manner. The designed apertures were then used to create a C-CASSI spectral source and to reconstruct the image sequence of interest. The performance of the designed apertures was compared with random colored-coded apertures (R-CCA) [1] with the same transmittance, with blue noise apertures (BNA) [19], and with 50% transmittance blocking-unblocking coded apertures (BUA) [21]. For the experiments, the estimated low-resolution image had a spatial resolution of 32×32 pixels, which corresponds to a spatial downsampling by a factor of 4. An example of the motion field obtained for the first frame, and the first spectral band is depicted in Fig. 2. Fig. 2(a) displays the motion field, Fig. 2(b) shows a zoomed version of the motion field, Fig. 2(c) depicts the Otzu thresholding of the motion field, which divides the scene into static and dynamic regions, Fig. 2(d) displays a hybrid blue noise coded aperture which is composed of one blue noise code aperture for the static part of the scene and a dynamic blue noise that moves to keep the complementarity between frames.

The quality of image reconstructions was evaluated in terms of peak-signal-to-noise ratio (PSNR) and structural similarity index (SSIM). The PSNR, given in decibels (dB), is related to the mean squared error (MSE) as $10\log_{10}(\max^2/\text{MSE})$ where $\max$ is the maximum possible value of an image pixel. SSIM measures the structure similarity between two imageswith values varying from 0 to 1, 1 being the value obtained for two identical images. Table I summarizes the results in terms of PSNR mean and SSIM for the different coded apertures. The PSNR and SSIM obtained with the proposed VA-CCA patterns are higher than those obtained with BUA, BNA, and R-CCA. Fig. 3 shows the RGB reconstructions of the frames $\#4$, $\#8$, and $\#12$ and provides the quality of reconstruc
of PSNR. The VA-CCA provides the best recons
a PSNR up to 2 dB higher than the block-u
apertures, 2.9 dB higher than blue noise coded
0.7 dB higher than the random colored-coded ap
results are very promising.

TABLE I
MEAN PSNR AND SSIM OF THE RECONSTRUCTION IN
SPECTRAL AND TEMPORAL DIMENSIONS

Coded Aperture Patterns	BU	BN	CCA
PSNR **mean**	25.04	24.56	26.75
SSIM **mean**	0.844	0.843	0.895

V. CONCLUSIONS

This paper studied a new design of adaptive
apertures (VA-CCA) for compressive spectra
video C-CASSI. After introducing a mathemat
dynamic scenes, we investigated the colored-c
in C-CASSI relying on prior information on t
vided by motion estimation between pairs
frames. The proposed design divides the vid
static and dynamic regions, assigning two different blue noise patterns to the static and mobile parts of the image. The improvement in PSNR obtained with the proposed approach is up to 2 dB compared to traditional blocking-unblocking apertures, 2.9 dB in comparison with non-adaptive blue noise patterns, and 0.7 dB higher than random colored-coded apertures.

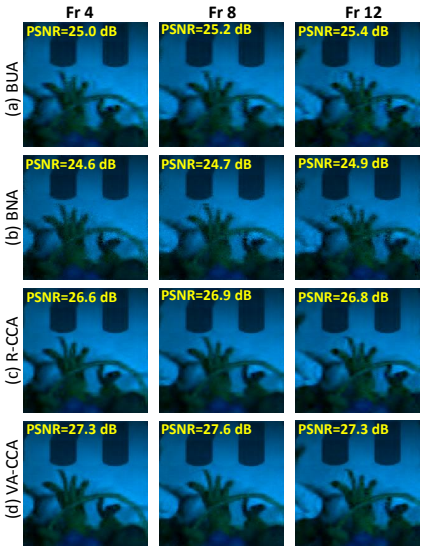
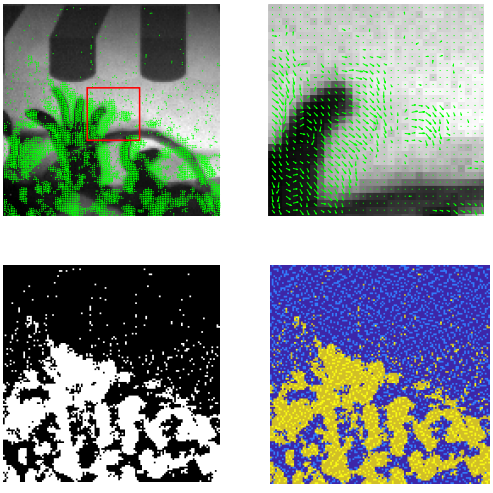


Fig. 3. Video reconstructions for frames 4, 8, and 12 (one frame per column), using non-adaptive and adaptive coded apertures. (a) block-unblock apertures, (b) blue noise apertures, (c) random CCA, (d) video adaptive CCA. The PSNR mean across the spectral band is shown in each frame.

ACKNOWLEDGMENTS

Part of this work was supported by the STIC-AmSud project 18-STIC-05. HyperMed - Image reconstruction from compressed measurements: application to hyperspectral and medical imaging.



(c) Threshold estimation (d) Adaptive Video CCA

Fig. 2. (a) Example of motion field of the 1st frame, and the 1st spectral band. (b) Zoomed version of the motion field which corresponds to the red square in (a). (c) Threshold motion estimation which divides in two the video scene, i.e., static area and dynamic area. (d) Hybrid blue noise coded aperture, notice that a yellow area depicts the dynamic blue noise coded aperture, and the blue area represent the static blue noise coded aperture.

REFERENCES

[1] H. Arguello and G. R. Arce, "Colored Coded Aperture Design by Concentration of Measure in Compressive Spectral Imaging," *IEEE Transactions on Image Processing*, vol. 23, no. 4, pp. 1896–1908, Apr. 2014.

[2] G. R. Arce, D. J. Brady, L. Carin, H. Arguello, and D. S. Kittle, "Compressive Coded Aperture Spectral Imaging: An Introduction," *IEEE Signal Processing Magazine*, vol. 31, no. 1, pp. 105–115, Jan. 2014.

[3] M. A. T. Figueiredo, R. D. Nowak, and S. J. Wright, "Gradient Projection for Sparse Reconstruction: Application to Compressed Sensing and Other Inverse Problems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 1, no. 4, pp. 586–597, Dec. 2007.

[4] A. Wagadarikar, R. John, R. Willett, and D. Brady, "Single disperser design for coded aperture snapshot spectral imaging," *Applied Optics*, vol. 47, no. 10, pp. B44–B51, Apr. 2008.

[5] K. M. León-López, L. V. G. Carreño, and H. A. Fuentes, "Temporal Colored Coded Aperture Design in Compressive Spectral Video Sensing," *IEEE Transactions on Image Processing*, vol. 28, no. 1, pp. 253–264, Jan. 2019.

[6] P. Llull, X. Liao, X. Yuan, J. Yang, D. Kittle, L. Carin, G. Sapiro, and D. J. Brady, "Coded aperture compressive temporal imaging," *Optics Express*, vol. 21, no. 9, pp. 10 526–10 545, May 2013.

[7] T.-H. Tsai, P. Llull, X. Yuan, L. Carin, and D. J. Brady, "Spectral-temporal compressive imaging," *Optics Letters*, vol. 40, no. 17, pp. 4054–4057, Sep. 2015.

[8] L. Wang, Z. Xiong, D. Gao, G. Shi, and a. F. Wu, "High-speed hyperspectral video acquisition with a dual-camera architecture," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2015, pp. 4942–4950.

[9] C. Ma, X. Cao, X. Tong, Q. Dai, and S. Lin, "Acquisition of High Spatial and Spectral Resolution Video with a Hybrid Camera System," *International Journal of Computer Vision*, vol. 110, no. 2, pp. 141–155, Nov. 2014.

[10] X. Cao, X. Tong, Q. Dai, and S. Lin, "High resolution multispectral video capture with a hybrid camera system," in *CVPR 2011*, Jun. 2011, pp. 297–304.

[11] N. Ouzir, A. Basarab, H. Liebgott, B. Harbaoui, and J. Tourneret, "Motion estimation in echocardiography using sparse representation and dictionary learning," *IEEE Transactions on Image Processing*, vol. 27, no. 1, pp. 64–77, Jan 2018.

[12] N. Ouzir, A. Basarab, O. Lairez, and J. Tourneret, "Robust optical flow estimation in cardiac ultrasound images using a sparse representation," *IEEE Transactions on Medical Imaging*, vol. 38, no. 3, pp. 741–752, March 2019.

[13] N. Diaz, A. Basarab, J. Yves, and H. Arguello, "Cardiac motion estimation using convolutional sparse coding," in *2019 27th European Signal Processing Conference (EUSIPCO), Coruña, España*, Sep 2019.

[14] N. Diaz, H. Rueda, and H. Arguello, "Adaptive filter design via a gradient thresholding algorithm for compressive spectral imaging," *Applied Optics*, vol. 57, no. 17, pp. 4890–4900, Jun. 2018.

[15] N. Diaz, C. Hinojosa, and H. Arguello, "Adaptive grayscale compressive spectral imaging using optimal blue noise coding patterns," *Optics & Laser Technology*, vol. 117, pp. 147–157, Sep. 2019.

[16] B. Wohlberg, "Efficient algorithms for convolutional sparse representations," *IEEE Transactions on Image Processing*, vol. 25, no. 1, pp. 301–315, Jan. 2016.

[17] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, Jan. 2011.

[18] B. Wohlberg, "Efficient algorithms for convolutional sparse representations," *IEEE Trans. Image Process.*, vol. 25, no. 1, pp. 301–315, Jan 2016.

[19] C. V. Correa, H. Arguello, and G. R. Arce, "Spatiotemporal blue noise coded aperture design for multi-shot compressive spectral imaging," *JOSA A*, vol. 33, no. 12, pp. 2312–2322, Dec. 2016.

[20] C. V. Correa-Pugliese, D. F. Galvis-Carreño, and H. Arguello, "Sparse representations of dynamic scenes for compressive spectral video sensing," *DYNA*, vol. 83, no. 195, pp. 42–51, Jan. 2016.

[21] D. F. Galvis-Carreño, Y. H. Mejía-Melgarejo, and H. Arguello-Fuentes, "Efficient reconstruction of Raman spectroscopy imaging based on compressive sensing," *DYNA*, vol. 81, no. 188, pp. 116–124, Dec. 2014.